

Perturbation bounds and degree of imprecision for uniquely convergent imprecise Markov chains

Damjan Škulj

University of Ljubljana, Faculty of Social Sciences
Kardeljeva pl. 5, SI-1000 Ljubljana, Slovenia
`damjan.skulj@fdv.uni-lj.si`

March 1, 2016

Abstract

The effect of perturbations of parameters for uniquely convergent imprecise Markov chains is studied. We provide the maximal distance between the distributions of original and perturbed chain and maximal degree of imprecision, given the imprecision of the initial distribution. The bounds on the errors and degrees of imprecision are found for the distributions at finite time steps, and for the stationary distributions as well.

Keywords. imprecise Markov chain, sensitivity of imprecise Markov chain, perturbation of imprecise Markov chains, degree of imprecision, weak coefficient of ergodicity

2010 Mathematics Subject Classification: 60J10

1 Introduction

Markov chains depend on a large number of parameters whose values are often subject to uncertainty. In the long run even small changes in the initial distribution or transition probabilities may cause large deviations. Several approaches to cope with uncertainty and estimate its magnitude have therefore been developed. Perturbation analysis gives estimates for the differences between probability distributions when the processes evolve in time ([9, 10, 14, 16]) or for the stationary distributions ([2, 15]) based on the differences in parameters (see also [12, 13] for the continuous time case).

In the recent decades variety of models of *imprecise probabilities* [1] have been developed. They provide means of expressing probabilistic uncertainty in the way that no reference to particular precise models is needed. The results thus reflect exactly the amount of uncertainty that results from the uncertainty in the inputs. Uncertainty in the parameters of Markov chains has first been addressed with models of this kind by Hartfiel [7], without formally connecting it to the theory of imprecise probabilities, although applying similar ideas. More recently, De Cooman *et al.* [4] formally linked Markov chains with the theory of upper previsions, while Škulj [19] proposed an approach based on the theory of interval probabilities. Unique convergence of imprecise Markov chains has been further

investigated by Hermans & De Cooman [8], where ergodicity of upper transition operators has been characterized in several ways, and by Škulj & Hable [21], who investigated the generalization of coefficients of ergodicity. Škulj [20] also studied the structure of non-uniquely convergent imprecise Markov chains.

Although at first glance models of imprecise probabilities and classical perturbation models seem to have the same objective, which they approach from different angles, this is often not so. In fact they answer essentially different questions. While imprecise probabilities provide models that replace classical probabilities with generalized models that are capable to reflect uncertainty or imprecision, perturbation models give the numerical information on how much uncertainty in results to expect, given the amount of uncertainty in inputs. Moreover, it is even quite natural to involve perturbation analysis to the models of imprecise probabilities.

The goal of this paper is thus to apply results from perturbation theory to imprecise Markov chains. There are two main reasons for this. The first is, that imprecise models too, are sensitive to the changes of input parameters. In the case of imprecise Markov chains this means that the changes in the bounds of input models affect the bounds of the distributions at further times in a similar way as in the case with the precise models. The other, maybe even more important reason is that currently in the theory of imprecise probabilities there has been little attention paid to the 'degree of imprecision', that is the maximal distances between, say, lower and upper bounds of probabilities. At least in comparison with the attention received by the methods for calculating these bounds. We thus also estimate how the 'degree of imprecision' evolves in time for imprecise Markov chains.

The paper is structured as follows. In Sc. 2 we introduce imprecise Markov chains with a special emphasis on the representation of the probability distributions after a number of time steps. In Sc. 3 we introduce the metric properties of imprecise operators, which allow us to measure the distances between imprecise probability distributions. By the means of the distances we also define the degree of imprecision. We show that with exception of the special case of 2-alternating upper probabilities (or 2-monotone lower probabilities) it is hard to find the exact distance between two imprecise probability models. In Sc. 4 we analyse the effects of perturbations of parameters to the deviations of the perturbed chains from the original ones. With a similar method we also study how the 'degree of imprecision' of the process grows in time. In Sc. 5 we apply the analysis to the case of contamination models and give a numerical example.

2 Imprecise Markov chains

2.1 Imprecise distributions and upper expectation functionals

Let \mathcal{X} be a finite set of *states*. A *probability distribution* of some random variable X over \mathcal{X} is given in terms of an *expectation functional* E on the space of real-valued maps f on \mathcal{X} , which we will denote by $\mathcal{L}(\mathcal{X})$:

$$E(f) = \sum_{x \in \mathcal{X}} P(X = x) f(x). \quad (1)$$

An *imprecise probability distribution* of a random variable X is given in terms of a closed convex set of expectation functionals \mathcal{M} , called a *credal set*. To every imprecise probability distribution a unique *upper expectation functional*

$$\overline{E}(f) = \sup_{E \in \mathcal{M}} E(f) \quad (2)$$

can be assigned. Upper expectation functionals are in a one-to-one correspondence with credal sets. Moreover, a functional \overline{E} is an upper expectation functional with respect to a credal set \mathcal{M} if and only if it satisfies the following properties:

- (i) $\min_{x \in \mathcal{X}} f(x) \leq \overline{E}(f) \leq \max_{x \in \mathcal{X}} f(x)$ (boundedness);
- (ii) $\overline{E}(f_1 + f_2) \leq \overline{E}(f_1) + \overline{E}(f_2)$ (subadditivity);
- (iii) $\overline{E}(\lambda f) = \lambda \overline{E}(f)$ (non-negative homogeneity);
- (iv) $\overline{E}(f + \mu 1_{\mathcal{X}}) = \overline{E}(f) + \mu$ (constant additivity);
- (v) if $f_1 \leq f_2$ then $\overline{E}(f_1) \leq \overline{E}(f_2)$ (monotonicity),

where $f, f_1, f_2 \in \mathcal{L}(\mathcal{X})$ are arbitrary, λ a non-negative real constant, μ an arbitrary real constant, and $1_{\mathcal{X}}$ the constant map 1 on \mathcal{X} . An upper expectation functional \overline{E} can be supplemented by a *lower expectation functional* with respect to the same credal set by assigning:

$$\underline{E}(f) = \min_{E \in \mathcal{M}} E(f). \quad (3)$$

The following duality relation holds:

$$\underline{E}(f) = -\overline{E}(-f) \quad (4)$$

for every $f \in \mathcal{L}(\mathcal{X})$. A lower expectation functional \underline{E} satisfies the same properties (i)–(v) as the upper ones, except for (ii), where subadditivity is replaced by

- (ii)' $\underline{E}(f_1 + f_2) \geq \underline{E}(f_1) + \underline{E}(f_2)$ (superadditivity).

2.2 Representations of uncertainty

In the previous section we described upper (and lower) expectation functionals, which uniquely represent convex sets of probability distributions. In principle such functionals possess properties similar to those of precise expectation functionals corresponding to precise probability distributions. However, there is a huge difference when it comes to the ways of specifying a convex set of probabilities compared to specifying a single probability distribution. In the latter case we need, in the general case, a single probability density function, or in the case finite spaces, a probability mass function. In the case of state spaces of finite Markov chains, we thus need to specify the probability of each state. In contrast, even in the case of finite spaces there are in general infinitely many values needed to specify a convex set of probability measures, or in the case of convex polytopes, this number is finite but often large.

Every convex polytope can be represented by specifying its extreme points or as an intersection of a set of half spaces separated by hyperplanes. The latter is far more natural and useful in the case of imprecise probabilities. This approach is behind many particular models, such as *lower* and *upper probabilities*, *probability intervals* or *coherent lower* and *upper previsions* [1, 11, 18]. The most general of those models are lower and upper previsions, which generalize all the other models, including lower and upper expectation functionals.

Although most researchers of imprecise probabilities list properties for lower previsions, in the theory of imprecise Markov chains, upper previsions are more often used. In general, an upper prevision $\overline{P}: \mathcal{K} \rightarrow \mathbb{R}$ is a map on some set \mathcal{K} – not necessarily a vector space – of measurable maps $\mathcal{X} \rightarrow \mathbb{R}$. What is important for our present model is the following equivalent definition of coherence for finite probability spaces.

Definition 1. Let $\mathcal{K} \subseteq \mathcal{L}(\mathcal{X})$ be a set of real-valued maps on a finite set \mathcal{X} and $\overline{P}: \mathcal{K} \rightarrow \mathbb{R}$ a mapping such that there exists a closed and convex set \mathcal{M} of (precise/linear) expectation functionals such that $\overline{P}(f) = \max_{P \in \mathcal{M}} P(f)$. Then \overline{P} is a coherent upper prevision on \mathcal{K} .

Note that any coherent upper prevision \overline{P} allows canonical extension to entire $\mathcal{L}(\mathcal{X})$ by defining

$$\overline{E}(f) = \max_{P \in \mathcal{M}} P(f). \quad (5)$$

This upper expectation functional is called the *natural extension* of \overline{P} , and is clearly itself too a coherent upper prevision on $\mathcal{L}(\mathcal{X})$. Upper expectation functionals thus form a subclass of coherent upper previsions.

Another subclass of imprecise probability models are *lower* and *upper probabilities*. A lower and upper probability \underline{P} and \overline{P} respectively are defined as a pair of real-valued maps on a class \mathcal{A} of subsets of \mathcal{X} . For every $A \in \mathcal{A}$, its probability is assumed to lie within $[\underline{P}(A), \overline{P}(A)]$. They too allow the formation of a credal set \mathcal{M} , whose members are exactly those expectation functionals E that satisfy the conditions $E(1_A) \geq \underline{P}(A)$ and $E(1_A) \leq \overline{P}(A)$ for every $A \in \mathcal{A}$. If the bounds $\underline{P}(A)$ and $\overline{P}(A)$ are reachable by the members of \mathcal{M} , then the lower/upper probabilities are said to be *coherent*.

When the lower and upper probabilities are defined on the set of elementary events $A = \{x\}$, where $x \in \mathcal{X}$, we are talking about *probability intervals (PRI)*, which, if coherent, also form a subclass of coherent upper previsions.

When an upper probability \overline{P} satisfies the following equation:

$$\overline{P}(A \cup B) + \overline{P}(A \cap B) \leq \overline{P}(A) + \overline{P}(B) \quad (6)$$

we call it *2-alternating*. At the same time, the corresponding lower probability \underline{P} , which satisfies equation $\underline{P}(A) = 1 - \overline{P}(A^c)$, is *2-monotone*:

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B). \quad (7)$$

If an imprecise probability model is given in the form of probability intervals $[p(x), \overline{p}(x)]$ for all elements $x \in \mathcal{X}$, we can first extend the bounds to all subsets of \mathcal{X} by

$$\underline{P}(A) = \max \left\{ \sum_{x \in A} p(x), 1 - \sum_{x \in A^c} \overline{p}(x) \right\}, \text{ for every } A \subseteq \mathcal{X}. \quad (8)$$

It can be shown that in this case \underline{P} is 2-monotone (see e.g. [3]).

2.3 Calculating the natural extension as a linear programming problem

Given a (coherent)¹ upper prevision \overline{P} on a set \mathcal{K} , its natural extension to $\mathcal{L}(\mathcal{X})$ can be calculated as a linear programming problem as follows:

Maximize

$$E(f) = \sum_{x \in \mathcal{X}} p(x)f(x) \quad (9)$$

subject to

$$\sum_{x \in \mathcal{X}} p(x) = 1 \quad (10)$$

$$\sum_{x \in \mathcal{X}} p(x)h(x) \leq \overline{P}(h) \quad (11)$$

for every $h \in \mathcal{K}$.

Although there exist efficient algorithms for solving linear programming problems, in the case of 2-monotone lower or 2-alternating upper probabilities the expectation bounds can be calculated even more efficiently by the use of *Choquet integral* (see e.g.[5]):

$$\overline{E}(f) = \min f + \int_{\min f}^{\max f} \overline{P}(f \geq x) dx \quad (12)$$

and

$$\underline{E}(f) = \min f + \int_{\min f}^{\max f} \underline{P}(f \geq x) dx \quad (13)$$

2.4 Imprecise transition operators

An (*imprecise*) *Markov chain* $\{X_n\}_{n \in \mathbb{N} \cup \{0\}}$ with the set of states \mathcal{X} is specified by an initial (imprecise) distribution and an (imprecise) transition operator. A *transition operator* assigns a probability distribution of X_{n+1} conditional on $(X_n = x)$. As before, we represent the conditional distribution with a conditional expectation functional $T(\cdot|x)$, mapping a real-valued map on the set of states into $T(f|x)$. A transition operator $T: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X})$ then maps $f \mapsto Tf$ such that

$$Tf(x) = T(f|x). \quad (14)$$

That is, the value of $Tf(x)$ equals the conditional expectation of f at time $n+1$ if the chain is in x at time n .

Replacing precise expectations $T(\cdot|x)$ with the imprecise ones given in terms of upper expectation functionals $\overline{T}(\cdot|x)$, we obtain an imprecise transition operator $\overline{T}: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{X})$ defined with

$$\overline{T}f(x) = \overline{T}(f|x). \quad (15)$$

¹Coherence is in fact not important for calculating the natural extension.

To every conditional upper expectation functional $\overline{T}(\cdot|x)$ a credal set can be assigned, and therefore a set of transition operators \mathcal{T} can be formed so that $\overline{T}f = \max_{T \in \mathcal{T}} Tf$.

Given an upper expectation functional \overline{E}_0 corresponding to the distribution of X_0 and an upper transition operator \overline{T} , we obtain

$$\overline{E}_n(f) = \overline{E}_0(\overline{T}^n f), \quad (16)$$

where \overline{E}_n is the upper expectation functional corresponding to the distribution of X_n .

Example 1. Let an imprecise transition operator be given in terms of a lower and upper transition matrices:

$$\underline{M} = \begin{bmatrix} 0.33 & 0.33 & 0 \\ 0.33 & 0.17 & 0.25 \\ 0 & 0.5 & 0.42 \end{bmatrix} \quad \text{and} \quad \overline{M} = \begin{bmatrix} 0.67 & 0.67 & 0 \\ 0.58 & 0.42 & 0.5 \\ 0 & 0.58 & 0.5 \end{bmatrix} \quad (17)$$

The upper transition operator is then

$$\overline{T}f = \max_{\substack{\underline{M} \leq M \leq \overline{M} \\ M1_{\mathcal{X}} = 1_{\mathcal{X}}}} Mf, \quad (18)$$

Thus, $\overline{T}f$ is the maximum of Mf over all row stochastic matrices that lie between \underline{M} and \overline{M} . Since each row can be maximized separately, the componentwise maximal vector does exist.

Further let

$$\underline{P}_0 = (0.33, 0.25, 0.25) \quad \text{and} \quad \overline{P}_0 = (0.38, 0.38, 0.42) \quad (19)$$

be the bounds for initial probability mass functions, whence the initial upper expectation functional is defined with

$$\overline{E}_0(f) = \max_{\substack{\underline{P}_0 \leq P \leq \overline{P}_0 \\ P(1_{\mathcal{X}}) = 1}} P(f). \quad (20)$$

Both \overline{T} and \overline{E}_0 are operators whose values are obtained as solutions of linear programming problems. As we explained in previous sections, we would hardly expect \overline{E}_n , that is the upper expectation functional for X_n , to be easily expressed in terms of a single linear programming problem, but rather as a sequence of problems. Thus, to obtain the value of $\overline{E}_n(f)$, for some $f \in \mathcal{L}(\mathcal{X})$, we would first find $\overline{T}f$ as a solution of the linear program (18) and then use it in the next instance of the linear program with the objective function replaced with the expectation of $\overline{T}f$ to obtain $\overline{T}^2 f$, until finally we would maximize $\overline{E}_0(\overline{T}^n f)$ as a linear program of the form (20). Practically this means that even though \overline{E}_0 is natural extension of a simple probability interval, the linear programs for \overline{E}_n are in general much more complex. A similar situation occurs with an n -step transition operator \overline{T}^n .

Nevertheless, we might still be interested in the lower bounds for probabilities of events of the form $(X_n = x)$. We can, for instance, find the upper probability $\overline{P}(X_n = x)$ as $\overline{E}_0 \overline{T}^n 1_{\{x\}}$, understanding of course that the imprecise probability model for the distribution of X_n is no longer a simple probability interval.

The above leads us to the idea, that even if the imprecise transition operators \overline{T}^n cannot be expressed in a simple form comparable to matrices, we might still want to provide the lower and upper transition matrix containing the information of the conditional probabilities $\underline{P}(X_{m+n} = y | X_m = x)$ and $\overline{P}(X_{m+n} = y | X_m = x)$, again bearing in mind that this is not an exhaustive information on the imprecise transition model. The upper probability can, for instance, be calculated by finding $\overline{T}^n 1_{\{y\}}(x)$.

In our case, the lower and upper probability mass vectors for X_3 are

$$\underline{P}_3 = (0.1966, 0.2672, 0.1513) \quad \text{and} \quad \overline{P}_3 = (0.5293, 0.5799, 0.3903), \quad (21)$$

and the lower and upper 3-step transition probabilities are

$$\underline{M}_3 = \begin{bmatrix} 0.2195 & 0.2500 & 0.1040 \\ 0.2195 & 0.2583 & 0.1533 \\ 0.1650 & 0.3067 & 0.2205 \end{bmatrix} \quad \text{and} \quad \overline{M}_3 = \begin{bmatrix} 0.5898 & 0.5992 & 0.3350 \\ 0.5383 & 0.5730 & 0.4175 \\ 0.4239 & 0.5609 & 0.4175 \end{bmatrix} \quad (22)$$

3 Metric properties of imprecise operators

3.1 Distances between upper operators

Let $\mathcal{L}_1 = \{f \in \mathcal{L}(\mathcal{X}) : 0 \leq f(x) \leq 1 \forall x \in \mathcal{X}\}$. In [21] the following distance between two upper expectation functionals \overline{E} and \overline{E}' is defined:

$$d(\overline{E}, \overline{E}') = \max_{f \in \mathcal{L}_1} |\overline{E}(f) - \overline{E}'(f)|. \quad (23)$$

When restricted to precise expectation functionals, the above distance coincides with the *total variation distance* for probability measures:

$$d(P, Q) = \max_{A \in \mathcal{F}} |P(A) - Q(A)|, \quad (24)$$

for two probability measures on an algebra \mathcal{F} . For real-valued maps on \mathcal{X} we use the *Chebyshev distance*

$$d(f, g) = \max_{x \in \mathcal{X}} |f(x) - g(x)|, \quad (25)$$

which we also extend to upper transition operators \overline{T} and \overline{T}' :

$$d(\overline{T}, \overline{T}') = \max_x d(\overline{T}(\cdot|x), \overline{T}'(\cdot|x)) = \max_{f \in \mathcal{L}_1} d(\overline{T}f, \overline{T}'f). \quad (26)$$

Subadditivity of the upper expectation functionals implies that

$$|\overline{E}(f) - \overline{E}(g)| \leq |\overline{E}(f - g)| \vee |\overline{E}(g - f)| \leq d(f, g). \quad (27)$$

Hence,

$$d(\overline{T}f, \overline{T}g) = \max_{x \in \mathcal{X}} |\overline{T}(f|x) - \overline{T}(g|x)| \leq d(f, g) \quad (28)$$

for every upper transition operator \overline{T} . Similarly, we have that

$$d(\overline{E}\overline{T}, \overline{E}\overline{T}') = \max_{f \in \mathcal{L}_1} |\overline{E}(\overline{T}f) - \overline{E}(\overline{T}'f)| \leq \max_{f \in \mathcal{L}_1} d(\overline{T}f, \overline{T}'f) = d(\overline{T}, \overline{T}'). \quad (29)$$

It is easy to see that the distances defined using the corresponding lower expectation functionals and transition operators are the same as those where the upper expectations are used.

In practical situations the upper expectations are usually the natural extensions of some coherent upper previsions defined on some subset $\mathcal{K} \subset \mathcal{L}(\mathcal{X})$. It would be therefore very useful if the differences between the values of two upper previsions on the elements of \mathcal{K} would give some information on the distances between their natural extensions. In general, however, this does not seem to be possible. For an illustration we give the following simple example.

Example 2. Let \mathcal{X} be a set of 3 states, say x, y, z and $\mathcal{K} = \{f_1 = (0, 1, 0), f_2 = (0.1, 1, 0)\}$. Then let two lower/upper previsions be given with the values:

$$\begin{array}{ll} \underline{P}_1(f_1) = 0.3 & \underline{P}_2(f_1) = 0.3 \\ \overline{P}_1(f_2) = 0.305 & \overline{P}_2(f_2) = 0.306 \end{array}$$

Now let \overline{E}_1 and \overline{E}_2 be the corresponding natural extensions and $h = (1, 0.5, 0)$. Then we have that

$$\overline{E}_1(h) = 0.2 \quad \text{and} \quad \overline{E}_2(h) = 0.21.$$

Thus, although the maximal distance between \overline{P}_1 and \overline{P}_2 on \mathcal{K} is only 0.001, the distance $d(\overline{E}_1, \overline{E}_2)$ is at least 0.01, which is 10 times larger.

Unlike the general case, in the case of 2-monotone lower probabilities the following holds ([21], Proposition 22):

Proposition 1. *Let \underline{P}_1 and \underline{P}_2 be 2-monotone lower probabilities and \underline{E}_1 and \underline{E}_2 the corresponding lower expectation functionals. Then*

$$d(\underline{E}_1, \underline{E}_2) = \max_{A \subseteq \mathcal{X}} |\underline{P}_1(A) - \underline{P}_2(A)|. \quad (30)$$

The fact that $\underline{P}_1(A) - \underline{P}_2(A) = (1 - \overline{P}_1(A^c)) - (1 - \overline{P}_2(A^c)) = \overline{P}_2(A^c) - \overline{P}_1(A^c)$ and $d(\overline{E}_1, \overline{E}_2) = d(\underline{E}_1, \underline{E}_2)$ implies that the choice of either upper or lower functionals does not make any difference.

3.2 Distances between upper and lower operators

Let \mathcal{M}_1 and \mathcal{M}_2 be two credal sets with the corresponding lower and upper expectation functionals denoted by $\overline{E}_1, \underline{E}_1$ and $\overline{E}_2, \underline{E}_2$ respectively. It has been shown in [21] that the maximal distance between the elements of two credal sets can be expressed in terms of the distance between the corresponding expectation functionals:

$$\overline{d}(\mathcal{M}_1, \mathcal{M}_2) := \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} d(E_1, E_2) \quad (31)$$

$$= \max_{f \in \mathcal{L}_1} \max\{\overline{E}_1(f) - \underline{E}_2(f), \overline{E}_2(f) - \underline{E}_1(f)\} \quad (32)$$

Now, since $1 - f \in \mathcal{L}_1$ iff $f \in \mathcal{L}_1$ and using $\overline{E}_i(1 - f) = 1 - \underline{E}_i(f)$, the above simplifies into:

$$= \max_{f \in \mathcal{L}_1} \overline{E}_1(f) - \underline{E}_2(f). \quad (33)$$

It follows that

$$\bar{d}(\mathcal{M}, \mathcal{M}) = \max_{f \in \mathcal{L}_1} \bar{E}(f) - \underline{E}(f), \quad (34)$$

which could be regarded as a measure of imprecision of a credal set. The above equalities justify the following definition of a distance between upper and lower expectation functionals.

$$d(\bar{E}_1, \underline{E}_2) = \max_{f \in \mathcal{L}_1} \bar{E}_1(f) - \underline{E}_2(f), \quad (35)$$

and

$$d(\bar{T}, \underline{T}) = \max_{x \in \mathcal{X}} d(\bar{T}(\cdot|x), \underline{T}(\cdot|x)). \quad (36)$$

The following proposition holds:

Proposition 2. *Let \bar{E}_1 and \underline{E}_2 be a lower and an upper expectation functionals. Then*

$$\max_{f \in \mathcal{L}_1} \bar{E}_1(f) - \underline{E}_2(f) = \max_{A \subseteq \mathcal{X}} \bar{E}_1(1_A) - \underline{E}_2(1_A), \quad (37)$$

where 1_A denotes the indicator function of set A . This implies that

$$d(\bar{E}_1, \underline{E}_2) = \max_{A \subseteq \mathcal{X}} \bar{E}_1(1_A) - \underline{E}_2(1_A). \quad (38)$$

Proof. Let \mathcal{M}_1 and \mathcal{M}_2 be the credal sets corresponding to \bar{E}_1 and \underline{E}_2 . We have that

$$d(\bar{E}_1, \underline{E}_2) = \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} d(E_1, E_2) \quad (39)$$

$$= \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} \max_{f \in \mathcal{L}_1} E_1(f) - E_2(f). \quad (40)$$

For every (precise) expectation functional E_i there exists some probability mass function p_i so that $E_i(f) = \sum_{x \in \mathcal{X}} p_i(x) f(x)$. Now let $A = \{x: p_1(x) \geq p_2(x)\}$ and let $F = 1_A$. For every $f \in \mathcal{L}_1$ we have that

$$E_1(f) - E_2(f) = \sum_{x \in \mathcal{X}} (p_1(x) - p_2(x)) f(x) \quad (41)$$

$$\leq \sum_{x \in \mathcal{X}} (p_1(x) - p_2(x)) F(x) \quad (42)$$

$$= E_1(F) - E_2(F). \quad (43)$$

Thus we have that the difference $E_1(f) - E_2(f)$ is always maximized by an indicator function. Hence,

$$d(\bar{E}_1, \underline{E}_2) = \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} \max_{A \subseteq \mathcal{X}} E_1(1_A) - E_2(1_A) \quad (44)$$

$$= \max_{A \subseteq \mathcal{X}} \left\{ \max_{E_1 \in \mathcal{M}_1} E_1(1_A) - \min_{E_2 \in \mathcal{M}_2} E_2(1_A) \right\} \quad (45)$$

$$= \max_{A \subseteq \mathcal{X}} \bar{E}_1(1_A) - \underline{E}_2(1_A). \quad (46)$$

□

3.3 Coefficients of ergodicity

Coefficients of ergodicity measure the rate of convergence of Markov chains. Given a metric d on the set of probability distributions, a coefficient of ergodicity is a real-valued map $\tau: T \mapsto \tau(T)$ with the property that

$$d(pT, qT) \leq \tau(T)d(p, q),$$

where p and q are arbitrary probability mass functions. In the general form coefficients of ergodicity were defined by Seneta [17], while in the form where the total variation distance is used, it was introduced by Dobrushin [6]. Given a stochastic matrix P , the value of $\tau(P)$ assuming the total variation distance, equals maximal distance between the rows of P :

$$\tau(P) = \max_{i,j} d(P_i, P_j), \quad (47)$$

where P_i and P_j are the i -th and j -th rows of P respectively. In the operators notation we would write

$$\tau(T) = \max_{x,y} d(T(\cdot|x), T(\cdot|y)). \quad (48)$$

The general definition clearly implies that:

$$d(pT^n, qT^n) \leq \tau(T)^n d(p, q). \quad (49)$$

Coefficients of ergodicity are also called *contraction coefficients*. Since transition operators are always non expanding, which means that $\tau(T) \leq 1$, the case of interest is usually when $\tau(T)$ is strictly less than 1. In such case $\tau(T)^n$ tends to 0 as n approaches infinity, which means that the distance $d(pT^n, qT^n)$ approaches 0. This means that the distance between probability distribution of random variables X_n of the corresponding Markov chain is diminishing, or equivalently, that the distributions converge to a unique limit distribution.

Often, despite $\tau(T) = 1$, the value of $\tau(T^r)$ might be strictly less than 1, which is sufficient to guarantee unique convergence. In fact, a chain is uniquely convergent exactly if $\tau(T^r) < 1$ for some positive integer r .

Coefficients of ergodicity have been generalized for the case of imprecise Markov chains too. The first, so called *uniform coefficient of ergodicity* has been introduced by Hartfiel [7] as

$$\tau(\mathcal{T}) = \max_{T \in \mathcal{T}} \tau(T), \quad (50)$$

where τ is the coefficient of ergodicity based on the total variation distance. If $\tau(\mathcal{T}) < 1$ this implies unique convergence of the corresponding Markov chains in the sense that every subset of the chains uniquely converges. This implies that the upper and lower expectations converge too, but in order to ensure unique convergence in the sense of expectation bounds weaker condition suffices.

The *weak coefficient of ergodicity* for imprecise Markov chains was defined in [21] as

$$\rho(\overline{T}) = \max_{x,y \in \mathcal{X}} d(\overline{T}(\cdot|x), \overline{T}(\cdot|y)). \quad (51)$$

That is, it is equal to the maximal distance between its row upper expectation functionals. The following properties hold:

- (i) $\rho(\overline{T}\overline{S}) \leq \rho(\overline{T})\rho(\overline{S})$ for arbitrary upper transition operators \overline{T} and \overline{S} ;
- (ii) $d(\overline{E}_1\overline{T}, \overline{E}_2\overline{T}) \leq d(\overline{E}_1, \overline{E}_2)\rho(\overline{T})$ for arbitrary upper expectation functionals $\overline{E}_1, \overline{E}_2$ and transition operator \overline{T} .

The following theorem holds.

Theorem 1 ([21] Theorem 21). *Let \overline{T} be an imprecise transition operator corresponding to a Markov chain $\{X_n\}_{n \in \mathbb{N}}$. Then the chain converges uniquely if and only if $\rho(\overline{T}^r) < 1$ for some integer $r > 0$.*

If \overline{T} is an upper transition operator such that $\overline{T}(\cdot|x)$ is the natural extension of some 2-alternating upper probability, then it follows from Proposition 1 that

$$\rho(\overline{T}) = \max_{A \subseteq \mathcal{X}} \max_{x, y \in \mathcal{X}} |\overline{T}(A|x) - \overline{T}(A|y)|. \quad (52)$$

Ergodicity coefficient can be applied to a pair of upper and lower expectation functionals as follows.

Proposition 3. *Let \overline{E}_1 and \underline{E}_2 be an upper and lower expectation functionals, and \overline{T} an upper transition operator. Then:*

$$d(\overline{E}_1\overline{T}, \underline{E}_2\overline{T}) \leq d(\overline{E}_1, \underline{E}_2)\rho(\overline{T}). \quad (53)$$

Proof. Denote by \mathcal{M}_1 and \mathcal{M}_2 the credal sets corresponding to \overline{E}_1 and \underline{E}_2 respectively. Then we have that:

$$d(\overline{E}_1\overline{T}, \underline{E}_2\overline{T}) = \max_{f \in \mathcal{L}_1} \overline{E}(\overline{T}f) - \underline{E}(\overline{T}f) \quad (54)$$

$$= \max_{f \in \mathcal{L}_1} \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} E_1(\overline{T}f) - E_2(\overline{T}f) \quad (55)$$

$$= \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} \max_{f \in \mathcal{L}_1} E_1(\overline{T}f) - E_2(\overline{T}f) \quad (56)$$

$$\leq \max_{E_1 \in \mathcal{M}_1, E_2 \in \mathcal{M}_2} d(E_1, E_2)\rho(\overline{T}) \quad (57)$$

$$= d(\overline{E}_1, \underline{E}_2)\rho(\overline{T}). \quad (58)$$

□

4 Perturbations of imprecise Markov chains

4.1 Distances between imprecise distributions of perturbed Markov chains

Suppose we have two imprecise Markov chains given by initial expectation functionals $\overline{E}_0, \overline{E}'_0$ and upper transition operators $\overline{T}, \overline{T}'$. The n -th step upper expectation functionals are then $\overline{E}_n = \overline{E}_0 \overline{T}^n$ and $\overline{E}'_n = \overline{E}'_0 \overline{T}'^n$ respectively. Our goal is to find the bounds on the distances between \overline{E}_n and \overline{E}'_n if the distances $d(\overline{E}_0, \overline{E}'_0)$ and $d(\overline{T}, \overline{T}')$ are known. We will also assume that both chains are uniquely convergent with weak coefficients of ergodicity $\rho_n = \rho(\overline{T}^n)$ and $\rho'_n = \rho(\overline{T}'^n)$, so that $\lim_{n \rightarrow \infty} \rho_n = 0$ and $\lim_{n \rightarrow \infty} \rho'_n = 0$. The latter conditions are clearly necessary and sufficient for unique convergence. Moreover, we will

give bounds on the distance between the limit distributions \overline{E}_∞ and \overline{E}'_∞ . To do so we will follow the similar derivation of the bounds for the case of precise (but not necessarily finite state) Markov chains by Mitrophanov [14].

We will make use of the following proposition:

Proposition 4. *The following equality holds for a pair of imprecise Markov chains and every $n \in \mathbb{N} \cup \{0\}$:*

$$\overline{E}_0 \overline{T}^n - \overline{E}'_0 \overline{T}'^n = (\overline{E}_0 \overline{T}^n - \overline{E}'_0 \overline{T}^n) + \sum_{i=0}^{n-1} (\overline{E}'_0 \overline{T}'^i \overline{T} \overline{T}^{n-i-1} - \overline{E}'_0 \overline{T}'^i \overline{T}' \overline{T}'^{n-i-1}), \quad (59)$$

and therefore,

$$d(\overline{E}_0 \overline{T}^n, \overline{E}'_0 \overline{T}'^n) \leq d(\overline{E}_0 \overline{T}^n, \overline{E}'_0 \overline{T}^n) + \sum_{i=0}^{n-1} d(\overline{E}'_0 \overline{T}'^i \overline{T} \overline{T}^{n-i-1}, \overline{E}'_0 \overline{T}'^i \overline{T}' \overline{T}'^{n-i-1}). \quad (60)$$

Theorem 2. *Denote $E_n = d(\overline{E}_n, \overline{E}'_n)$ and $D = d(\overline{T}, \overline{T}')$. The following inequality holds:*

$$E_n \leq E_0 \rho_n + D \sum_{i=0}^{n-1} \rho_i. \quad (61)$$

Proof. Proposition 4 implies that

$$E_n \leq d(\overline{E}_0 \overline{T}^n, \overline{E}'_0 \overline{T}'^n) + \sum_{i=0}^{n-1} d(\overline{E}'_0 \overline{T}'^i \overline{T} \overline{T}^{n-i-1}, \overline{E}'_0 \overline{T}'^i \overline{T}' \overline{T}'^{n-i-1}) \quad (62)$$

denote $\overline{E}'_{i+1} = \overline{E}'_0 \overline{T}'^{i+1}$ and $\overline{E}_{i+1}^* = \overline{E}'_0 \overline{T}'^i \overline{T}$

$$\leq E_0 \rho_n + \sum_{i=0}^{n-1} d(\overline{E}_{i+1}^* \overline{T}^{n-i-1}, \overline{E}'_{i+1} \overline{T}'^{n-i-1}) \quad (63)$$

$$\leq E_0 \rho_n + \sum_{i=0}^{n-1} d(\overline{E}_{i+1}^*, \overline{E}'_{i+1}) \rho_{n-i-1} \quad (64)$$

$$= E_0 \rho_n + \sum_{i=0}^{n-1} d(\overline{E}'_i \overline{T}, \overline{E}'_i \overline{T}') \rho_{n-i-1} \quad (65)$$

by (29)

$$\leq E_0 \rho_n + \sum_{i=0}^{n-1} d(\overline{T}, \overline{T}') \rho_{n-i-1} \quad (66)$$

$$= E_0 \rho_n + D \sum_{i=0}^{n-1} \rho_{n-i-1} \quad (67)$$

$$= E_0 \rho_n + D \sum_{i=0}^{n-1} \rho_i. \quad (68)$$

□

Corollary 1. Denote $D_n = d(\overline{T}^n, \overline{T}'^n)$ (that is $D_1 = D$). The following inequality then holds:

$$D_n \leq D_1 \sum_{i=0}^{n-1} \rho_i. \quad (69)$$

Proof. We have:

$$d(\overline{T}^n, \overline{T}'^n) = \max_{f \in \mathcal{L}_1} d(\overline{T}^n f, \overline{T}'^n f) \quad (70)$$

$$= \max_{f \in \mathcal{L}_1} \max_{x \in \mathcal{X}} d(\overline{T}(\overline{T}^{n-1} f | x), \overline{T}'(\overline{T}'^{n-1} f | x)) \quad (71)$$

$$= \max_{f \in \mathcal{L}_1} \max_{x \in \mathcal{X}} d(\overline{T}(\cdot | x) \overline{T}^{n-1} f, \overline{T}'(\cdot | x) \overline{T}'^{n-1} f) \quad (72)$$

$$= \max_{x \in \mathcal{X}} d(\overline{T}(\cdot | x) \overline{T}^{n-1}, \overline{T}'(\cdot | x) \overline{T}'^{n-1}) \quad (73)$$

by Theorem 2

$$\leq \max_{x \in \mathcal{X}} d(\overline{T}(\cdot | x), \overline{T}'(\cdot | x)) \rho_{n-1} + D_1 \sum_{i=0}^{n-2} \rho_i \quad (74)$$

$$= D_1 \rho_{n-1} + D_1 \sum_{i=0}^{n-2} \rho_i = D_1 \sum_{i=0}^{n-1} \rho_i. \quad (75)$$

□

Lemma 1. Let \overline{T} be an upper transition operator such that $\rho(\overline{T}^r) = \rho_r =: \rho < 1$ and let $n = kr + m$, where $m < r$. Then

$$\sum_{i=0}^{n-1} \rho_i \leq r \frac{1 - \rho^k}{1 - \rho} + m \rho^k \quad (76)$$

and

$$\sum_{i=0}^{\infty} \rho_i \leq \frac{r}{1 - \rho}. \quad (77)$$

Proof. Clearly $\{\rho_n\}$ is a non-increasing sequence. Moreover, it follows directly from the definitions and monotonicity that

$$\rho_i \leq \rho^{r \lceil \frac{i}{r} \rceil}, \quad (78)$$

where $\lceil \cdot \rceil$ denotes the integer part. The required equations are now obtained by taking sums of the left and right hand sides of equation (78). □

Let E_∞ denote the distance between the limit distributions \overline{E}_∞ and \overline{E}'_∞ . The following corollary is a direct consequence of the above results.

Corollary 2. Using the notation from Theorem 2 and Lemma 1 we have the following inequalities

$$E_n \leq E_0 \rho^k + D \left(r \frac{1 - \rho^k}{1 - \rho} + m \rho^k \right) \quad (79)$$

$$E_\infty \leq \frac{Dr}{1 - \rho} \quad (80)$$

and

$$D_n \leq D_1 \left(r \frac{1 - \rho^k}{1 - \rho} + m \rho^k \right). \quad (81)$$

Remark 1. Notice that $D_\infty = E_\infty$ for every uniquely convergent Markov chain.

4.2 Degree of imprecision

Let $\{X_n\}_{n \in \mathbb{N}}$ be an imprecise Markov chain and let \overline{E}_n denote the upper expectation functionals corresponding to the imprecise distributions of X_n . As a measure of the degree of imprecision, we have suggested in Sc. 3.2 $I_n = d(\overline{E}_n, \underline{E}_n)$. Our goal in this section is to find bounds on I_n given the initial imprecision I_0 and the imprecision of the transition operator, given by $\hat{I} = d(\overline{T}, \underline{T})$.

Similarly as in Proposition 4 we have the following.

Proposition 5. *Let \overline{E}_0 and \underline{E}_0 be a pair of lower and upper expectation functionals and \overline{T} and \underline{T} an upper and lower transition operators. Then we have that*

$$\overline{E}_0 \overline{T}^n - \underline{E}_0 \underline{T}^n = \overline{E}_0 \overline{T}^n - \underline{E}_0 \overline{T}^n + \sum_{i=0}^{n-1} (\underline{E}_0 \underline{T}^i \overline{T} \overline{T}^{n-i-1} - \underline{E}_0 \underline{T}^i \underline{T} \overline{T}^{n-i-1}), \quad (82)$$

and therefore

$$d(\overline{E}_0 \overline{T}^n, \underline{E}_0 \underline{T}^n) = d(\overline{E}_0 \overline{T}^n, \underline{E}_0 \overline{T}^n) + \sum_{i=0}^{n-1} d(\underline{E}_0 \underline{T}^i \overline{T} \overline{T}^{n-i-1}, \underline{E}_0 \underline{T}^i \underline{T} \overline{T}^{n-i-1}). \quad (83)$$

Theorem 3. *Let $I_n = d(\overline{E}_n, \underline{E}_n)$, $\hat{I} = d(\overline{T}, \underline{T})$ and $\rho_n = \rho(\overline{T}^n)$. The following inequality holds:*

$$I_n \leq I_0 \rho_n + \hat{I} \sum_{i=0}^{n-1} \rho_i. \quad (84)$$

Proof. It follows directly from Proposition 3 that

$$d(\overline{E}_0 \overline{T}^n, \underline{E}_0 \overline{T}^n) \leq I_0 \rho_n. \quad (85)$$

Further we have that for every $0 \leq i \leq n-1$

$$d(\underline{E}_0 \underline{T}^i \overline{T} \overline{T}^{n-i-1}, \underline{E}_0 \underline{T}^i \underline{T} \overline{T}^{n-i-1}) = \max_{f \in \mathcal{L}_1} |\underline{E}_0[\underline{T}^i \overline{T} \overline{T}^{n-i-1} f] - \underline{E}_0[\underline{T}^i \underline{T} \overline{T}^{n-i-1} f]| \quad (86)$$

which is by definition, and by replacing $\underline{E}_i = \underline{E}_0 \underline{T}^i$ we have

$$= \max_{f \in \mathcal{L}_1} |\underline{E}_i[\overline{T} \overline{T}^{n-i-1} f] - \underline{E}_i[\underline{T} \overline{T}^{n-i-1} f]| \quad (87)$$

by (27)

$$= \max_{f \in \mathcal{L}_1} d(\overline{T} \overline{T}^{n-i-1} f, \underline{T} \overline{T}^{n-i-1} f) \quad (88)$$

$$= \max_{f \in \mathcal{L}_1} \max_{x \in \mathcal{X}} d(\overline{T}(\overline{T}^{n-i-1} f|x), \underline{T}(\overline{T}^{n-i-1} f|x)) \quad (89)$$

$$= \max_{x \in \mathcal{X}} d(\overline{T}(\cdot|x) \underline{T}^{n-i-1}, \underline{T}(\cdot|x) \underline{T}^{n-i-1}) \quad (90)$$

by Proposition 3

$$= \max_{x \in \mathcal{X}} d(\overline{T}(\cdot|x), \underline{T}(\cdot|x)) \rho_{n-i-1} \quad (91)$$

$$= \hat{I} \rho_{n-i-1}. \quad (92)$$

Now the required inequality follows directly by combining the above inequalities. \square

The following corollaries now follow immediately using similar reasoning as in the case of distances in Sc. 4.1.

Corollary 3. *Let \overline{T} be an upper transition operator and denote $\hat{I}_n = d(\overline{T}^n, \underline{T}^n)$. The following inequality holds:*

$$\hat{I}_n \leq \hat{I}_1 \sum_{i=0}^{n-1} \rho_i.$$

Corollary 4. *Using the notation from Theorem 3 and Lemma 1 we have the following inequalities*

$$I_n \leq I_0 \rho^k + \hat{I} \left(r \frac{1 - \rho^k}{1 - \rho} + m \rho^k \right) \quad (93)$$

$$I_\infty \leq \frac{\hat{I} r}{1 - \rho} \quad (94)$$

and

$$\hat{I}_n \leq \hat{I}_1 \left(r \frac{1 - \rho^k}{1 - \rho} + m \rho^k \right). \quad (95)$$

Remark 2. It is again clear that for every uniquely convergent imprecise Markov chain $\hat{I}_\infty = I_\infty$.

5 Examples

5.1 Contamination models

Let \overline{E} be an upper expectation functional and $\varepsilon > 0$. Then we consider the ε -contaminated upper expectation functional

$$\overline{E}_\varepsilon(f) = (1 - \varepsilon) \overline{E}(f) + \varepsilon f_{\max}, \quad (96)$$

where $f_{\max} = \max_{x \in \mathcal{X}} f(x) =: \bar{V}(f)$. Note that \bar{V} is the upper expectation functional whose credal set consists of all expectation functionals on $\mathcal{L}(\mathcal{X})$. It is called the *vacuous upper prevision*. The upper transition operator $\bar{T}_V(f) = f_{\max} 1_{\mathcal{X}}$ is called the *vacuous upper transition operator*.

Being a convex combination of \bar{E} and the vacuous upper prevision \bar{V} , \bar{E}_ε is itself also an upper prevision. Similarly we could define an ε -contaminated upper transition operator with

$$\bar{T}_\varepsilon f = (1 - \varepsilon)\bar{T}f + \varepsilon 1_{\mathcal{X}} f_{\max} = (1 - \varepsilon)\bar{T}f + \varepsilon \bar{T}_V f. \quad (97)$$

Let $\rho = \rho(\bar{T})$. Then we can explicitly find the coefficients of ergodicity for the contaminated model.

Proposition 6. *Let \bar{E} be an upper expectation functional, \bar{T} an upper transition operator and \bar{E}_ε and \bar{T}_ε the corresponding ε -contaminated models. The following inequalities hold:*

- (i) $d(\bar{E}, \bar{E}_\varepsilon) = \varepsilon d(\bar{E}, \bar{V})$, where \bar{V} is the vacuous upper prevision;
- (ii) $d(\bar{T}, \bar{T}_\varepsilon) = \varepsilon d(\bar{T}, \bar{T}_V)$, where \bar{T}_V is the vacuous upper transition operator;
- (iii) $d(\bar{E}'_\varepsilon, \bar{E}_\varepsilon) = (1 - \varepsilon)d(\bar{E}', \bar{E})$;
- (iv) $d(\bar{T}_\varepsilon, \bar{T}'_\varepsilon) = (1 - \varepsilon)d(\bar{T}, \bar{T}')$;
- (v) $\rho(\bar{T}_\varepsilon) = (1 - \varepsilon)\rho(\bar{T})$;
- (vi) $d(\bar{E}_\varepsilon, \underline{E}_\varepsilon) = (1 - \varepsilon)d(\bar{E}, \underline{E}) + \varepsilon$;
- (vii) $\hat{I}(\bar{T}_\varepsilon) = (1 - \varepsilon)\hat{I}(\bar{T}) + \varepsilon$.

Proof. (i) follows directly from

$$d(\bar{E}, \bar{E}_\varepsilon) = \max_{f \in \mathcal{L}_1} |\bar{E}(f) - (1 - \varepsilon)\bar{E}(f) - \varepsilon \bar{V}(f)| \quad (98)$$

$$= \max_{f \in \mathcal{L}_1} |\varepsilon(\bar{E}(f) - \bar{V}(f))| \quad (99)$$

$$= \varepsilon d(\bar{E}, \bar{V}) \quad (100)$$

and simply leads to (ii). To see (iii) we calculate

$$d(\bar{E}_\varepsilon, \bar{E}'_\varepsilon) = \max_{f \in \mathcal{L}_1} |(1 - \varepsilon)\bar{E}(f) + \varepsilon \bar{V}(f) - (1 - \varepsilon)\bar{E}'(f) - \varepsilon \bar{V}(f)| \quad (101)$$

$$= \max_{f \in \mathcal{L}_1} |\varepsilon(\bar{E}(f) - \bar{E}'(f))| \quad (102)$$

$$= \varepsilon d(\bar{E}, \bar{E}'). \quad (103)$$

(iv) and (v) are direct consequences of (iii) and the definitions. To see (vi) note that $\underline{V}(f) = f_{\min}$ and therefore $\underline{E}_\varepsilon(f) = (1 - \varepsilon)\underline{E}(f) + \varepsilon \underline{V}(f)$. For some f we then have

$$\bar{E}_\varepsilon(f) - \underline{E}_\varepsilon(f) = (1 - \varepsilon)(\bar{E}(f) - \underline{E}(f)) + \varepsilon(f_{\max} - f_{\min}) \quad (104)$$

Now suppose the maximal difference in the above expression is attained for some $f \in \mathcal{L}_1$ and denote $\tilde{f} = \frac{f - f_{\min}}{f_{\max} - f_{\min}}$ which belongs to \mathcal{L}_1 as well. It is directly

verified that $\overline{E}(\tilde{f}) - \underline{E}(\tilde{f}) = \frac{1}{f_{\max} - f_{\min}}(\overline{E}(f) - \underline{E}(f))$. Thus, $f_{\max} - f_{\min} = 1$ must hold, because of maximality of f , and therefore (vi) easily follows.

(vii) is also a simple consequence of (vi). \square

Theorem 4. *Let \overline{E} and \overline{T} be an upper expectation functional and an upper transition operator respectively, and \overline{E}_ε and \overline{T}_ε the corresponding ε -contaminated operators. Denote $E_n = d(\overline{E}_{\varepsilon n}, \underline{E}_n)$, where $\overline{E}_{\varepsilon n} = \overline{E}_\varepsilon \overline{T}_\varepsilon^n$ and $\underline{E}_n = \underline{E} \overline{T}^n$, $\Delta_1 = d(\overline{E}, \underline{V})$, $\Delta_2 = d(\overline{T}, \underline{T}_V)$, $\hat{I} = \hat{I}(\overline{T})$, $\rho = \rho(\overline{T})$ and the imprecision of the contaminated chain by $I'_n = d(\overline{E}_{\varepsilon n}, \underline{E}_{\varepsilon n})$.*

Then we have that $E_0 = \varepsilon \Delta_1$, $D = \varepsilon \Delta_2$ and $\hat{I}(\overline{T}_\varepsilon) = (1 - \varepsilon)\hat{I}(\overline{T}) + \varepsilon$. The following inequalities hold:

$$E_n \leq \varepsilon \Delta_1 \rho^n (1 - \varepsilon)^n + \varepsilon \Delta_2 \frac{1 - \rho^n (1 - \varepsilon)^n}{1 - \rho(1 - \varepsilon)}; \quad (105)$$

$$E_\infty \leq \frac{\varepsilon \Delta_2}{1 - \rho(1 - \varepsilon)}; \quad (106)$$

$$I'_n \leq ((1 - \varepsilon)I_0 + \varepsilon)(1 - \varepsilon)^n \rho^n + ((1 - \varepsilon)\hat{I} + \varepsilon) \frac{1 - \rho^n (1 - \varepsilon)^n}{1 - \rho(1 - \varepsilon)}; \quad (107)$$

$$I'_\infty \leq \frac{(1 - \varepsilon)\hat{I} + \varepsilon}{1 - \rho(1 - \varepsilon)}. \quad (108)$$

Proof. A direct consequence of Theorems 2 and 3, Proposition 6 and Corollary 2. \square

5.2 Numerical example

We again consider a Markov chain with the initial lower and upper probabilities and transition probabilities as in Example 1. We compare it with a perturbed chain whose lower and upper transition matrices are

$$\underline{M}' = \begin{bmatrix} 0.32 & 0.36 & 0 \\ 0.36 & 0.19 & 0.24 \\ 0 & 0.5 & 0.4 \end{bmatrix} \quad \text{and} \quad \overline{M}' = \begin{bmatrix} 0.64 & 0.68 & 0 \\ 0.57 & 0.38 & 0.45 \\ 0.04 & 0.56 & 0.46 \end{bmatrix} \quad (109)$$

and the initial probability bounds are

$$\underline{P}'_0 = (0.32, 0.21, 0.28) \quad \text{and} \quad \overline{P}'_0 = (0.42, 0.38, 0.42) \quad (110)$$

Coefficients of ergodicity are $\rho(\overline{T}) = 0.67$ and $\rho(\overline{T}') = 0.60$, and the distance between initial imprecise probability models is $d(\overline{E}_0, \overline{E}'_0) = 0.0248$, and between transition operators $d(\overline{T}, \overline{T}') = 0.05$.

The maximal theoretically possible bounds $d(E_n, E'_n)$ can be obtained using Theorem 2, with $E_0 = 0.0248$, $D = 0.05$ and $\rho_n = \rho(\overline{T}')^n = 0.60^n$. For comparison we have calculated lower and upper transition probability matrices and the distances based on these estimates. The results are listed in Figure 1. The actual distances may be larger because the expectation functionals are not fully described by probability interval models (PRI).

In Figure 2 the distances between the operators \overline{T}^n and \overline{T}'^n are given, together with their upper bounds calculated using Corollary 1.

Figure 1: Distances between \overline{E}_n and \overline{E}'_n based on PRI estimates and their theoretical upper bounds.

n	1	2	3	∞
PRI-distance	0.0248	0.0387	0.0429	0.0467
maximal theoretical distance	0.0740	0.0889	0.1034	0.1250

Figure 2: Distances between \overline{T}^n and \overline{T}'^n based on PRI estimates and their theoretical upper bounds.

n	2	3	4	∞
PRI-distance	0.0454	0.0499	0.0484	0.0467
maximal theoretical distance	0.0800	0.0980	0.1088	0.1250

6 Conclusions and further work

We have studied the impact of perturbations of initial imprecise probability distributions and transition operators of imprecise Markov chains on the deviations of distributions of the chain at further steps. The results show that stability of the distributions depends on the weak coefficient of ergodicity, which is consistent with the known results for precise Markov chains [14]. By the same means we give the bounds on the degree of imprecision depending on the imprecision of initial distribution and transition operators.

Our goal in the future is to extend the results to related models, such as continuous time imprecise Markov chains, hidden Markov models or semi-Markov models.

Acknowledgement

The author would like to thank Alexander Y. Mitrophanov for his helpful suggestions and discussions.

References

- [1] Augustin, Thomas, Coolen, Frank PA, de Cooman, Gert, & Troffaes, Matthias CM. 2014. *Introduction to imprecise probabilities*. John Wiley & Sons.
- [2] Cho, Grace E, & Meyer, Carl D. 2001. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications*, **335**(1), 137–150.
- [3] De Campos, Luis M, Huete, Juan F, & Moral, Serafin. 1994. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **2**(02), 167–196.

- [4] De Cooman, Gert, Hermans, Filip, & Quaeghebeur, Erik. 2009. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, **23**(04), 597–635.
- [5] Denneberg, Dieter. 1994. *Non-additive measure and integral*. Vol. 27. Springer Science & Business Media.
- [6] Dobrushin, Roland L. 1956. Central limit theorem for nonstationary Markov chains. I, II. *Theory of Probability & Its Applications*, **1**(1), 65–80.
- [7] Hartfiel, Darald J. 1998. *Markov Set-Chains*. Berlin: Springer-Verlag.
- [8] Hermans, Filip, & De Cooman, Gert. 2012. Characterisation of ergodic upper transition operators. *International Journal of Approximate Reasoning*, **53**(4), 573–583.
- [9] Kartashov, NV. 1986a. Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. *Theory of Probability & Its Applications*, **30**(2), 247–259.
- [10] Kartashov, NV. 1986b. Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. II. *Theory of Probability & Its Applications*, **30**(3), 507–515.
- [11] Miranda, Enrique, & De Cooman, Gert. 2007. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, **46**(1), 188–225.
- [12] Mitrophanov, Alexander Y. 2003. Stability and exponential convergence of continuous-time Markov chains. *Journal of applied probability*, **40**(4), 970–979.
- [13] Mitrophanov, Alexander Y. 2005a. Ergodicity coefficient and perturbation bounds for continuous-time Markov chains. *Mathematical Inequalities and Applications*, **8**(1), 159–168.
- [14] Mitrophanov, Alexander Y. 2005b. Sensitivity and convergence of uniformly ergodic Markov chains. *Journal of Applied Probability*, 1003–1014.
- [15] Mouhoubi, Zahir, & Aïssani, Djamil. 2010. New perturbation bounds for denumerable Markov chains. *Linear Algebra and its Applications*, **432**(7), 1627–1649.
- [16] Rudolf, Daniel, & Schweizer, Nikolaus. 2015. Perturbation theory for Markov chains via Wasserstein distance. *arXiv preprint arXiv:1503.04123*.
- [17] Seneta, Eugene. 1979. Coefficients of ergodicity: structure and applications. *Advances in applied probability*, 576–590.
- [18] Troffaes, Matthias CM, & De Cooman, Gert. 2014. *Lower previsions*. John Wiley & Sons.
- [19] Škulj, Damjan. 2009. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, **50**(8), 1314–1329.

- [20] Škulj, Damjan. 2013. A classification of invariant distributions and convergence of imprecise Markov chains. *Linear Algebra and its Applications*, **439**(9), 2542 – 2561.
- [21] Škulj, Damjan, & Hable, Robert. 2013. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, **76**(1), 107–133.